

Crowdsourcing for a Geographical and Social Mapping of Italian Dialects

François Bry
Institute for Informatics
Ludwig-Maximilian
University
Munich, Germany
bry@lmu.de

Fabian Kneissl
Institute for Informatics
Ludwig-Maximilian
University
Munich, Germany
fabian.kneissl@ifi.lmu.de

Thomas Krefeld
Institute for Romanic
Philology
Ludwig-Maximilian
University
Munich, Germany
thomas.krefeld@lmu.de

Stephan Lücke
IT Group Humanities
Ludwig-Maximilian
University
Munich, Germany
luecke@lmu.de

Christoph Wieser
Institute for Informatics
Ludwig-Maximilian
University
Munich, Germany
christoph.wieser@ifi.lmu.de

ABSTRACT

Linguistic field research depends on collecting phrases and sentences as well as their geographical and social characteristics. The traditional method of field research –researchers asking questions and filling forms– is time-consuming, costly, and not free of biases. This article presents *metropolitalia*, a Web-based crowdsourcing platform for linguistic field research aiming at overcoming some of the drawbacks of traditional linguistic field research. *metropolitalia* is built upon *Agora*, a market for trading with phrases and speculating on their characteristics in a playful manner. Two games are run under *Agora*, *Borsa Parole* and *Poker Parole*, that aim at collecting complementary data and meta-data: *Borsa Parole* incites players to express their own knowledge or, rather, beliefs, *Poker Parole* incites players to make conjectures on the beliefs of others, thus enhancing the primary meta-data collected with *Borsa Parole* with secondary, or reflexive, meta-data needed for language perception studies. This article describes *Agora* with both games and reports on first evaluations of the data gathered so far.

Author Keywords

Crowdsourcing, Games With A Purpose (GWAP), Linguistic Field Research

ACM Classification Keywords

H.3.5 Information Storage and Retrieval: Online Information Services—*Web-based services*; J.5 Arts and Humanities: Linguistics

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WebSci'13, May 1 – May 5, 2013, Paris, France.
ACM 978-1-4503-1889-1.

INTRODUCTION

Linguistic field research is concerned with gathering and analysing speech data from speakers of some language(s) under observation. The data gathered comprise the speech data itself as well as characteristics of the speakers such as their geographical location and social characteristics, like age, gender, or level of education. Traditionally, such multi-dimensional data are collected by sending scientists, typically doctoral students or other low paid researchers, to the speakers' locations, usually in certain geographical regions, where they interview speakers, record and/or transliterate the interview, and report on these interviews by filling forms. This process is time-consuming because each researcher can only interview a limited number of speakers, costly because the researchers or students involved have to be paid, and furthermore can be biased because of (conscious or unconscious) preconceptions an interviewer might have [6]. As a consequence, only relatively limited areas can be covered by traditional linguistic field research.

The crowdsourcing platform *metropolitalia* –accessible at <http://www.metropolitalia.org> since August 2012– is conceived as a Web-based platform for linguistic field research [11]. It encourages people to participate in the process of gathering a large linguistic dataset from a wide geographical area with low costs for the linguists. Such a participation of many users to reach certain goals –that are not necessarily known to the users– is called crowdsourcing, a current trend on the Web which provides a cost- and time-efficient way of gathering data [7]. One way to gather data using crowdsourcing is by employing games known as “games with a purpose” (“GWAP”) [21], which is the option we describe in this article.

We designed two market-based games, both run under the same system *Agora* (Greek for “market”) for data gathering, on which symbolic goods can be traded and speculated with. On games based on *Agora*, people can submit symbolic goods

–like dialect phrases– together with their own assessment of characteristics of that symbolic good –where or within which social group the dialect phrase is used– and compare their own assessments with those of the community. Thus, one can speculate in both senses of forming conjectures and investing money with a symbolic good and its characteristics. One then receives a payment in form of both points, which can be seen as play-money or as tokens of expertise, when the community agrees. Agora is first used as operating system of a game called *Borsa Parole*, Italian for “word stock exchange”. On *Borsa Parole*, the better phrases and their characteristics are recognized by the user community, the more successful is a user expressing the same belief. Thus, mainly phrases with widely acknowledged linguistic traits are gathered with *Borsa Parole*. A demonstration of *Borsa Parole*’s most important aspects is shown in a screencast available at <http://www.vimeo.com/59723042>. A second game called *Poker Parole*, Italian for “word poker”, is accessible since March 2013 and also run with Agora, gathering complementary data. Here in contrast, forming conjectures on phrases’ characteristics that are only recognized by few people leads to successful play-money and reputation investments.

The Italian language is especially interesting for linguistic field research, making *Borsa Parole* and *Poker Parole* excellent means for investigating how to perform linguistic field research via crowdsourcing. Indeed, the Italian language spoken today everywhere, in cities and countryside alike, and within all social groups is currently undergoing a divergence that originates in the big cities and spreads from there [12]. This makes today’s Italian different from languages such as German, English, or French.

During the restructuring and standardisation process which the Italian language experienced only in the late 19th century, that is more recently than most other European languages, a common language emerged out of several rather disparate dialects. However, instead of being perceived as languages for less educated people, the Italian vernaculars –that is, unstandardized language varieties– and dialects –that is, languages socially or geographically subordinate to a (national or regional) standard language– have remained in today’s spoken and written language across all social groups [13]. A witness of the strength of the Italian dialects is their presence on Wikipedia: There are small but lively versions of Wikipedia in about a dozen of Italian dialects. Currently, the vernaculars spoken in large Italian cities evolve. Especially, new vernaculars emerge, gradually dissociating the metropolises from one another [12].

The difficulty of traditional linguistics to gather geographically and socially diverse data is especially salient with the Italian vernaculars and dialects. The manifold vernaculars differ from (standardised) dialects and from each other in vocabulary, grammar, and/or pronunciation. Some distinctive features in language use are well known in whole, or major parts of, Italy like for example the use of “bon dì” for greeting in some valleys in South Tyrol, others are mainly used by certain social groups like “delizioso” (meaning “cute”) mainly by women [12]. Other distinctive features in language use

are, in contrast, known only in limited parts of Italy. For linguistic research, the rarely recognized phrases are just as important as the well recognized phrases. So far, there is not much data available concerning Italian vernaculars and dialects. Thus, the platform *metropolitalia* described in this article are likely to gain much importance.

In essence, this article demonstrates how linguistic field research can be performed by Web-based crowdsourcing. Agora accounts for this need by providing the exploitation systems for two games for gathering quantitative data and different kinds of data which complement each other – as with the two complementary games *Borsa Parole* and *Poker Parole*. A similar approach can be imagined for other languages than Italian, though it must be adapted to the specifics of the language.

The contributions of this article are as follows:

- Presentation of the market-like operating system Agora.
- Presentation of two games, *Borsa Parole* and *Poker Parole*, both run by Agora, both aiming at gathering complementary linguistic data and meta-data.
- First evaluation of data gathered with *Borsa Parole*

RELATED WORK

The research reported about in this article is related to crowdsourcing in linguistics, “games with a purpose” (GWAP), and prediction markets.

Crowdsourcing denotes the participation of many humans on the Web to achieve a common goal [7]. Crowdsourcing is applied in many different contexts, like the collaborative web platforms Wikipedia and Yahoo! Answers or games solving image labeling tasks. Also in linguistics, crowdsourcing has already been applied successfully, mainly in theoretical linguistics. Munro et al. present in [17] linguistic projects exploiting human computation, specifically, Amazon Mechanical Turk (AMT), where users are paid for completing tasks. An important conclusion of this article is that the linguistic quality achieved using human computation is comparable to that of controlled laboratory studies. The majority of linguistic research relies on mechanised labour, like that AMT provides, for gathering data [19]. For example, Arabic dialects have been gathered via AMT to improve machine translation [23]. Further articles report on using GWAP for gathering corpora annotations [18, 9]. Duolingo¹ is a platform offering its users support in learning languages while collecting material for automated text translation. Furthermore, passive, observation-based approaches to analyzing social media for linguistics are investigated. For example, geotagged Twitter messages are gathered, automatically categorized into topics, and the geographical distribution of all terms measured, resulting in a geographical mapping of certain dialect terms [8].

Similar to crowdsourcing, human computation refers to applications, in which humans consciously or unconsciously collaborate to solve problems that so far can not be solved purely algorithmically [14]. If a game is designed such that

¹<http://www.duolingo.com>

users solve this problem while playing the game, the application is called a GWAP [21]. Von Ahn and Dabbish have introduced the term GWAP with the ESP Game that solves the image labeling problem. Here, the same image is shown to two randomly paired users who are rewarded if they suggest the same label for that image. Since the only resource shared by the two users is the image, the users tend to enter descriptions that are likely to be given also by their counterparty user. Thus, images are labeled with descriptions while users are playing the game. Also in art history, the GWAP on the “Artigo” platform are employed to gather descriptive tags for artworks [20]. Suggestions for an extension of the ESP Game are given in [3]. Several other GWAP have been designed that solve different problems, among others a game for protein-folding [5].

Prediction markets are employed for estimating what the results of unknown future events are. In prediction markets, users trade contracts whose payoff depends on unknown future events [22]. The idea is that in an efficient market, the price of such a contract directly correlates with the probability of the future event. Prediction markets are supposed to be efficient markets, which has been confirmed by research, and therefore can quite closely predict future events. For example, prediction markets are successful in elections and also outperform polls impressively [1]. Note that some researchers have expressed the view that direct estimates might be more precise than those generated on a prediction market [16].

To the best authors’ knowledge, no other crowdsourcing using games than Borsa Parole and Poker Parole have been proposed so far that rely on a market for gathering data for linguistic field research.

AGORA: A MARKET FOR GATHERING DATA

Agora is a generic software for running Web-based playmarkets in which a community of users can share symbolic goods as well as assessments of characteristics of these symbolic goods. A symbolic good can be a text (as in metropolitalia), an image, an audio file, or any other immaterial good (or combination thereof) that should be characterized by users. The good is symbolic in the sense that it can occur on Agora multiple times, be possessed by multiple users, and –technically– be transferable over the Internet. Agora makes it possible for a user to:

- add her own symbolic goods to the market,
- propose assessments for her own symbolic goods as well as for symbolic goods proposed by others,
- review and adapt her own assessments based on assessments from other users, and
- trade assessments with other users.

As depicted in Figure 1, an assessment consists of a user assessing one or more characteristics of a symbolic good and additionally estimating which proportion of users are likely to assign the same characteristics as she does. All assessments for a symbolic good together represent the market’s view for the symbolic good and if a user agrees with the aggregated

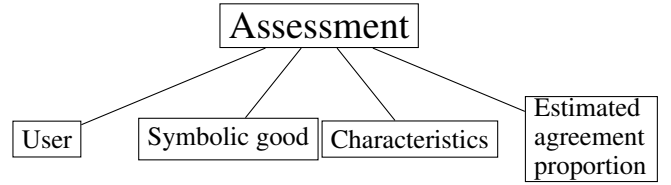


Figure 1. Composition of an assessment: A user assesses the characteristics of a symbolic good together with her estimated proportion of users agreeing on the characteristics.

view of the market, she gains (play-)money. The closer her estimation is to the proportion of users assigning the same characteristic (=agreement), the more money she gains. Assessments can be offered for sale for a user-defined price and bought by other users. Thus users can create their own portfolio of assessments and gather assessments they deem to be important or valuable.

To compute the monetary value of an assessment, the real agreement of other users to a user’s assessment needs to be calculated. From the real agreement and the user’s agreement, the monetary value can be computed. A real agreement value of 1 means that all other users agree with the user’s assessment, a value of 0 means that no other users agree.

DEFINITION 1. *Given a user u , a symbolic good g , a characteristic c , and an estimated agreement proportion p , an **assessment** is the tuple $a = (u, g, c, p)$. The set of all assessments is A .*

DEFINITION 2. *Given an assessment $a = (u, g, c, p)$, the **agreement** : $A \rightarrow [0, 1]$ is defined as:*

$$\text{agreement}(a) = \frac{\sum_{a_i \in A_{g,c}} s(a_i, a)}{|A_{g,c}|}$$

where

- $A_{g,c} \subseteq A$: all assessments on symbolic good g with the same type of characteristics as c ,
- $s(a_1, a_2) : A \times A \rightarrow [0, 1]$: function representing the similarity between the two assessments, and
- $|\cdot|$: cardinality of the set.

The function s for calculating similarity between two assessments has to be adapted to the needs of the specific system. It can be the Kronecker delta function whose result for a specified symbolic good would be 1 if the assessments have the same characteristics and 0 otherwise. But also elaborate similarity functions can be defined depending on the current context.

The calculated agreement can then be compared to the estimated agreement proportion of the assessment to yield the monetary value of an assessment. The smaller the difference, the higher the monetary value. A simple, linear function for an assessment is the following.

DEFINITION 3. *Given an assessment $a = (u, g, c, p)$, the linear function based monetary value of this assessment is:*

$$\text{value}_{\text{linear}}(a) = 100 \cdot (1 - |\text{agreement}(a) - p|)$$

where $|v|$ denotes the absolute value of v .

By multiplying a number like 100 to the pure difference the value is more accessible to users than decimal numbers between 0 and 1. Also other functions to define the monetary value are possible, e.g., the density function of a normal distribution which values close estimations higher and remote estimations lower than a linear function like the one given above, thus promoting good estimations.

DEFINITION 4. Given an assessment $a = (u, g, c, p)$, the normal distribution based monetary value of this assessment is:

$$value_{nd}(a) = 100 \cdot e^{-\frac{(agreement(a)-p)^2}{2\sigma^2}}$$

where σ^2 is the variance of the normal distribution.

For example, if σ is set to $\frac{1}{3}$, the range of values is the same as in $value_{linear}$, only the distribution is different (as shown in Figure 2).²

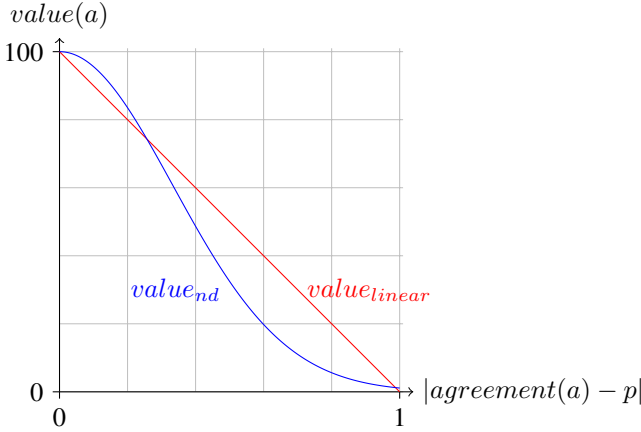


Figure 2. Two functions $value_{linear}$ and $value_{nd}$ with $\sigma = \frac{1}{3}$ expressing the monetary value of an assessment. The x-axis is the difference of the estimated agreement proportion and the real agreement.

If over time the agreement of an assessment diverges from the user's estimation, the user loses a part of the money the assessment was worth before. If it converges to her estimation, she gains money. By submitting an assessment, a user cannot lose money gained through other assessments. When a user reconsiders her assessments, for each one a summary of the other users' assessments is displayed. Based on this feedback she can adjust her assessments to fit the market. Here, the market regulates itself and users are rewarded for visiting the platform again. As in real markets, rules can be defined to limit the amount or frequency of changes of an estimation, e.g., through imposing a transaction cost for each change.

In order to effectively gather data with social media operated by Agora, users are encouraged to suggest symbolic goods themselves. This is important to enliven the media run on Agora so that they can grow both in the number of symbolic goods gathered and in the number of their users.

²To be exact, 99.73% of the values of $value_{nd}$ are in the range. Thus if values are given as integers without decimal, the ranges of both value functions are the same.

The market of Agora is similar to a financial market like Wall Street in that users speculate on characteristics of (symbolic) goods. In a financial market, participants buy and sell company shares which have a monetary value that represents the company's value. In Agora characteristics for symbolic goods are the financial market's shares. The estimated agreement proportion of such an assessment can be seen as the target price which a user wants to reach with her assessment, because upon reaching that agreement proportion she gains the most money from her assessment. The monetary value of an assessment represents the current price her assessment is worth. While users assess symbolic goods—similar to trading in financial markets—the values of the involved assessments of other users vary. As effort for buying an assessment the user has to play instead of spending money in order to motivate users to play games built upon Agora and therefore contribute data.

Agora differs from a financial market as follows: Agora is a play-market, that is, no real money is involved. Furthermore, the ownerships of symbolic goods, characteristics and assessments are symbolic in the sense that several users can “own” the same “good” (i.e., assessment on the same symbolic good with the same characteristics and estimated agreement proportion).

Specifically on the platform metropolitalia, Agora is used for running two games, Borsa Parole and Poker Parole, where Italian dialect or vernacular phrases—that is sentences or parts of sentences—are “traded with”.³ In other possible applications of Agora, completely different symbolic goods could be traded with, as discussed in section . Except in that section, the symbolic goods meant in this article are phrases in Italian dialects or vernaculars.

BORSA PAROLE: TRADING WITH ONE'S OWN BELIEFS

The goals of Borsa Parole are to gather new phrases and to encourage users to share their own assessments on new or existing phrases. Specifically, the user is asked to indicate where a phrase is spoken, how many people recognize the phrase as being from that location, which word(s) of the phrase are linguistically distinct, and who the speakers are in terms of age, gender, and level of education. For that purpose, three web pages exist that correspond to the three possible user actions (without trading assessments) of Agora, one for adding new phrases, one for assessing existing phrases, and one for reviewing and adapting own assessments. The trade of assessments is excluded in this first version of Borsa Parole for the sake of simplicity and will be added at a later stage. In the following, we focus on the assessment as it is the most interesting action. Borsa Parole is played in several rounds (we experiment with 3 rounds which we encountered as a good number also for casual players). Each round, one phrase is presented to the user which the user has to assess. The following can be done by the user one after the other:

³So far, the game provides written sentences but an extension with spoken sentences is foreseen. This extension does not require any change in the media logic but only additional user interfaces for collecting and rendering spoken language.



Figure 3. Borsa Parole during the choice of a region for the displayed sentence. The currently selected region (northern Italy) is highlighted in blue.

- choosing the geographical area where the phrase is spoken (see Figure 3),
- specifying her belief how many other users assign the same region,
- selecting individual words of the phrase that guided the user's geographical mapping,
- characterizing social attributes of speakers of the phrase.

For choosing a region, the user interface provides a top-down approach –stepwise focusing on smaller regions: broad geographical regions, political regions, provinces, and municipalities– as well as a bottom-up approach –an input field with automatic suggestions of regions.

Each user action is optional, i.e., can be skipped, to give the user freedom in her choice and to prevent false data if a user does not know what to choose. After all rounds, a summary is shown in which the user can see how other users characterized the phrases.

For being successful on Borsa Parole, one has to submit phrases with characteristics that many other users of Borsa Parole are likely to agree with, because there it is easier for others to determine the characteristics. As a consequence, success on Borsa Parole depends on how one is skilled at

forecasting others' conceptions. This is a typical case of a "beauty contest", as Keynes described the effect in a speculative market where participants reflect on each others' behaviour and adapt their behaviour accordingly [10]. However, while the beauty contest analogy was meant by Keynes as a criticism of speculation on financial markets, a beauty contest-like speculation contributes to the aim of Borsa Parole. Indeed, in linguistic field research the true opinion of a single speaker is much less relevant than her perception of the community's opinion. In other words, for linguistic field research, speculating speakers are welcome!

POKER PAROLE: SPECULATION ON THE BELIEFS OF OTHERS

Poker Parole, also being based on Agora, shares many properties of its game-play with Borsa Parole, with one exception: While success on Borsa Parole comes from submitting commonly recognized phrases, on Poker Parole it comes from submitting phrases that most users are not likely to recognize. Such phrases are equally important for linguistic research and therefore need to be gathered as well. The two games therefore complement each other in the data they gather.

The user is asked to give a phrase with a characterization that is hardly known by anybody living outside the chosen geographical or social area. So the speculation consists in telling

the community: “I guess that most of you won’t be capable to correctly recognize the characteristics of the following sentence.” Thus, users performing well in Poker Parole must be specialists for niche vernaculars or dialects, opposed to users performing well in Borsa Parole who must be generalists for widely known vernaculars or dialects.

INCENTIVES: THE FUEL OF GWAP

Borsa Parole and Poker Parole –and GWAP in general– can only gather much and manifold data if they provide enough incentives for users to engage in the games. Besides the game being designed for generating high quality data, providing the right and enough incentives is the main factor for success of a GWAP. Therefore, also the GWAP Borsa Parole and Poker Parole provide incentives which are described in the following.

First, the design as game that entertains and motivates the user is an incentive in itself. This also includes the gaming aspects (play-)money / points, highscore lists, and game rounds as further incentives. To avoid the user’s discouragement, she can skip phrases or characterizations she does not know or want to give.

Research suggests that user incentives can be most effective when incorporating the positive *social facilitation* effect and avoiding the negative *social loafing* [15]. Social facilitation describes that users tend to solve simple tasks better with someone else watching them than without supervision, while social loafing describes that users make less effort to solve tasks when working in a group than alone. Thus, the accomplishments of individual users should be shown prominently –highscore lists show the top performing users–, other users should be able to evaluate each user’s contribution –all entered characteristics are displayed in the results view for a phrase–, and the unique value of each user’s contribution should be highlighted –a summary of played rounds is shown to the user after each game highlighting her own actions. By incorporating such social psychological incentives, users tend to contribute more data and return to the Web platform [4].

Also, performing well on a market is an incentive in itself. This is true for financial markets like Wall Street as well as play-markets with symbolic goods. People’s interest in performing better than the crowd is apparent in both types of markets, although in financial markets prospects of earning money play a role as incentive as well. Furthermore, each kind of market involves a gaming dimension in itself as traders are playing with each other with their speculations in order to get the best performance on the market. These incentives also are apparent in the market-like games Borsa Parole and Poker Parole, where the (play-)money the user has depends on her speculation as well as on the other users’ assessments.

Concerning language, in all cultures there is a considerable interest in language issues and in reflecting on one’s own language variations. People interested in their own language are likely to participate in Borsa Parole and Poker Parole just for seeing what others disclose on the platform, both phrases or sentences they do not know and assessments they are not

aware of. Also if a user were not attracted by games in general but interested in language variations, she might still consider playing the games for the sake of her interest in language.

BEYOND ITALIAN LINGUISTICS

Agora is designed as a generic and modular system and therefore its deployment also in other application areas than Italian linguistics is possible.

For example in the area of art history, a similar application of the two complementary GWAP would yield new insights into the perception of artworks. The social goods traded with would be artworks and the characteristics assessed could be the artist, style, and epoch. Other than changing the GUI appropriately for displaying images of artworks instead of phrases and choosing the characteristics appropriately, the software for running such artwork-oriented games would stay the same, that is, Agora.

Also in other areas where there are both general and expert knowledge the complementarity of media in the style of Borsa Parole and Poker Parole are likely to be exploitable.

FIRST EVALUATION OF DATA GATHERED WITH BORSA PAROLE

For this evaluation, data was gathered on the platform metropolitalia with the game Borsa Parole during the first seven months of its public availability (from August 2012 until February 2013). No additional incentives, such as money, were rewarded to users other than those described above in section . For attracting users, personal contacts of the authors were informed, blog articles were published at Italian Web blogs, a blog accompanying the platform was established, and a Facebook site was setup.

This evaluation is divided into two parts: data quantity, giving results regarding user motivation and acceptance of the market-based game play, and data quality, showing that the data gathered is useful for research of Italian dialects.

Data Quantity

During the seven months period, 595 users visited the platform for playing 3530 rounds of Borsa Parole in total. Within these, 2121 times a geographical location was assigned by the user, 1959 times a geographical assessment (that is, a geographical characterization with estimated agreement proportions) was created, 1726 times one or several words were highlighted as being relevant, and 1037 social characterizations produced. The numbers show that 40% of all rounds were skipped, probably because the user did not know the phrase well enough to estimate a geographic region of its occurrence. This is natural and was foreseen, giving users the option to skip rounds. The high number of geographical assessments compared to geographical assignments without an estimation on the agreement proportion (92%) indicates that users are confident in giving such estimations, a finding that encourages to employ the market-based approach Agora in further games. The decreasing numbers of word selections and social characterizations bear evidence that completing

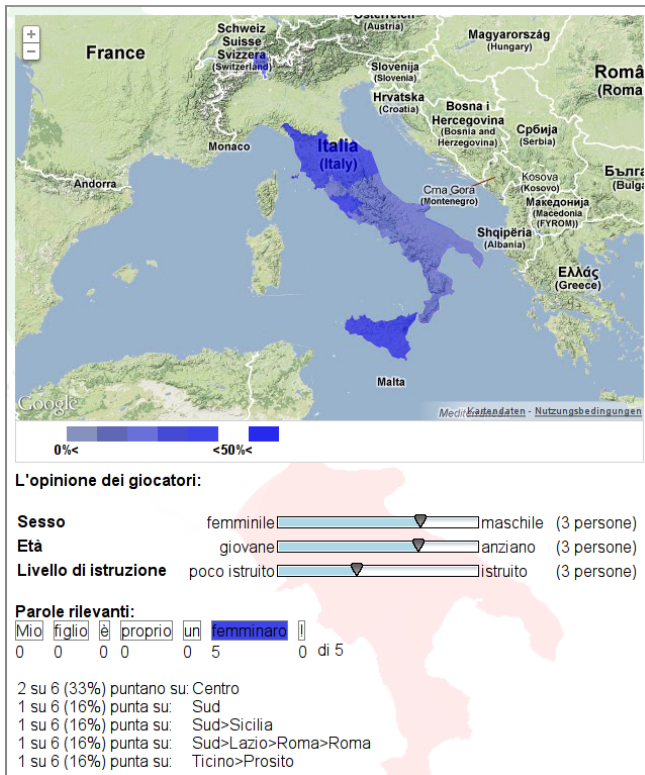


Figure 4. metropolitalia platform displaying the data gathered for the sentence “Mio figlio è proprio un femminaro!” (in English: “My son really is a womanizer!”)

these steps are optional. Furthermore for many phrases a social characterization can not be given because it does not exist from a linguistic point of view.

Not only the data gathering process on the characterization side can be seen to be successful, also the possibility for users to add new phrases or sentences led to 112 new phrases that were contributed to metropolitalia by users. Thus 11% of all users who played Borsa Parole at least once contributed new phrases. This indicates that the incentives for adding phrases to Borsa Parole are good enough. Compared to the number of all users visiting the platform metropolitalia, the percentage of phrase contributing users is 0.03%. This is on par with the contribution percentage of users on other social media sites, e.g., Wikipedia estimates 0.02-0.03%.⁴

Data Quality

Also the quality of the data gathered is convincing. In Figure 4, the results for a phrase as displayed on the platform are shown. The phrase is assessed to be spoken more in the south of Italy (see the coloured map), the speaker characterized as male, older, and less educated (see the three sliders), and the selected relevant word is “femminaro”, a vernacular word for a womaniser. Though only six users assessed the phrase so far, a clear tendency to the use in the center and south of Italy can be seen. And according to a native Italian

⁴Wikipedia estimates that 0.02-0.03% of all visitors actively contribute to Wikipedia. Source: http://strategy.wikimedia.org/wiki/Wikimedia_users?oldid=44715

Geographical region	Estim. a.p.	Real a.p.	Value
North	27%	50%	12
North	81%	50%	2
North	100%	50%	0
North>Laives	34%	45%	56
North>Lombardia	44%	50%	86
North>Rimini	10%	45%	1
South>Calabria	45%	13%	2
South>Matera	51%	24%	5
South>Taranto	35%	18%	31

Table 1. Gathered geographical assessments for the phrase “Se non la smetti ti do una sberla.” (in English “If you don’t stop it I’ll box your ears.”) including the estimated and the real agreement proportion and the monetary value of that assessment.

speaker knowing this word, it is well known in Sicily (island in the south of Italy).

The estimated agreement proportions of phrases are of mixed quality. In table 1, the geographical assessments for an exemplary phrase “Se non la smetti ti do una sberla.” (in English “If you don’t stop it I’ll box your ears.”) are shown. The word “sberla” (in English “slap in the face”) originally spread from northern Italy [2]. Some assessments are quite precise (as the ones for the northern Italian regions Lombardia and Laives) and are worth a lot, while others do not estimate the agreement of users on the platform well.

CONCLUSION

For linguistic field research, crowdsourcing has the potential to gather a huge amount of data from many people in a cost-effective way. The approach furthermore lowers the risk of biased data as data is directly entered into the platform by the speakers themselves without interpretation or other processing by interviewers. And also the variety of users can lessen biased results. As another advantage, crowdsourcing improves on traditional linguistic field research in the possibility to conduct long-term studies over several years. It is comparably inexpensive to run a Web-based platform for several years. To benefit from these advantages, the crowdsourcing platform has to be designed to gather the data wanted in the quality needed and at the same time attract enough users to reach its objectives. Both goals are fulfilled by metropolitalia and the two GWAP Borsa Parole and Poker Parole, both built upon the market-based system Agora. The market-based game design provides new incentives for users which in the evaluation is indicated to be accepted by users. Its analogy to speculation on real markets furthermore yields richer meta-data for evaluation than traditional questionnaire-based field research. A future evaluation of Poker Parole will give further insights into the users’ acceptance of the complementarity of the two GWAP.

ACKNOWLEDGMENTS

We thank Hubertus Kohle from the Institute for Art History at the Ludwig-Maximilian University of Munich and all other Play4Science project members for useful suggestions.

This research has been funded in part by the German Foundation of Research (DFG) within the project Play4Science

REFERENCES

1. Berg, J., Nelson, F., and Rietz, T. Prediction Market Accuracy in the Long Run. *International Journal of Forecasting* 24, 2 (2008), 283–298.
2. Bertinetto, P. M. La sillabazione dei nessi /sC/ in italiano: Un'eccezione alla tendenza 'universale'? In *Atti de XXXI Congresso Internazionale di Studi della Società di Linguistica Italiana, "Fonologia e morfologia dell'italiano e dei suoi dialetti"* (1999). (In Italian).
3. Bry, F., and Wieser, C. Squaring and Scripting the ESP Game: Trimming a GWAP to Deep Semantics. In *Proceedings of the International Conference on Serious Games Development and Applications* (2012).
4. Cheshire, C., and Antin, J. The Social Psychological Effects of Feedback on the Production of Internet Information Pools. *Computer-Mediated Communication* 13, 3 (2008), 705–727.
5. Cooper, S., Treuille, A., Barbero, J., Fay, A. L., Tuite, K., Khatib, F., Snyder, A. C., Beenen, M., Salesin, D., Baker, D., and Popović, Z. The Challenge of Designing Scientific Discovery Games. In *Proceedings of the 5th International Conference on the Foundation of Digital Games (FDG)* (2010).
6. Davis, K. A. Qualitative Theory and Methods in Applied Linguistics Research. *TESOL Quarterly* 29, 3 (1995), 427–453.
7. Doan, A., Ramakrishnan, R., and Halevy, A. Y. Crowdsourcing Systems on the World-Wide Web. *Communications of the ACM* 54, 4 (2011), 86–96.
8. Eisenstein, J., Connor, B. O., Smith, N. A., and Xing, E. P. A Latent Variable Model for Geographic Lexical Variation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2010).
9. Hladká, B., Mírovský, J., and Schlesinger, P. Designing a Language Game for Collecting Coreference Annotation. In *Proceedings of the 3rd Linguistic Annotation Workshop (ACL-IJCNLP)* (2009).
10. Keynes, J. M. *The General Theory of Employment, Interest, and Money*. Macmillan Cambridge University Press, 1936.
11. Kneissl, F., and Bry, F. Metropolitalia: A Crowdsourcing Platform for Linguistic Field Research. In *Proceedings of the IADIS International Conference WWW/Internet* (2012).
12. Krefeld, T. Italienische Varietätenlinguistik. *Italienisch. Zeitschrift für italienische Sprache und Literatur* 63 (2010), 56–62. (In German).
13. Krefeld, T. Alter Standard - Neue Medien. Zur Erfassung von Restandardisierungsprozessen im Italienischen. In *Koineisierung und Standardisierung in der Romania*, S. Dessi Schmid, J. Hafner, and S. Heinemann, Eds. Universitätsverlag Winter, 2011. (In German).
14. Law, E., and von Ahn, L. Human Computation. In *Synthesis Lectures on Artificial Intelligence and Machine Learning*, R. J. Brachman, W. W. Cohen, and T. Dietterich, Eds. Morgan & Claypool Publishers, 2011, 1–121.
15. Liu, Y., Lehdonvirta, V., Alexandrova, T., and Nakajima, T. Drawing on Mobile Crowds via Social Media. *Multimedia Systems* 18, 1 (2012), 53–56.
16. Manski, C. F. Interpreting the Predictions of Prediction Markets. Tech. rep., National Bureau of Economic Research, 2004.
17. Munro, R., Bethard, S., Kuperman, V., Lai, V. T., Melnick, R., Potts, C., Schnoebelen, T., and Tily, H. Crowdsourcing and Language Studies: The new Generation of Linguistic Data. In *Proceedings of the Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk at NAACL-HLT* (2010).
18. Poesio, M., Chamberlain, J., Kruschwitz, U., Robaldo, L., and Ducceschi, L. Phrase Detectives: Utilizing Collective Intelligence for Internet-Scale Language Resource Creation. *ACM Transactions on Interactive Intelligent Systems* (2012).
19. Sabou, M., Bontcheva, K., and Scharl, A. Crowdsourcing Research Opportunities: Lessons from Natural Language Processing. In *Processing of the 12th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW)* (2012).
20. Steinmayr, B., Wieser, C., Kneissl, F., and Bry, F. Karido: A GWAP for Telling Artworks Apart. In *Proceedings of the 16th International Conference on Computer Games (CGAMES)* (2011).
21. von Ahn, L., and Dabbish, L. Labeling Images with a Computer Game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (2004).
22. Wolfers, J., and Zitzewitz, E. Prediction Markets. *Journal of Economic Perspectives* 18, 2 (2004), 107–126.
23. Zbib, R., Malchiodi, E., Devlin, J., Stallard, D., Matsoukas, S., Schwartz, R., Makhoul, J., Zaidan, O. F., and Callison-Burch, C. Machine Translation of Arabic Dialects. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics* (2012).